

# Phraseografie und Korpusanalyse

Sören Stumpf (Trier)

---

## Abstract

The following article gives an overview of the weak points in the lexicographical coverage of phrasemes. The main problem with previous phraseography is that the dictionary entries are not based on comprehensive corpus analyses of actual language use. Hence I make a case for a “corpus-based phraseography” (Steyer 2010) and in using selected examples, I demonstrate how a pragmatic approach that is focused on actual language use can help to improve the lemmatization of formulaic expressions. This also shows which consequences and changes may occur from a corpus-analytical point of view as compared to the traditional phraseographical approach. For this purpose, I use the German reference corpus/Deutsches Referenzkorpus and the analysis system COSMAS II. Central to my analysis are the phenomena that have scarcely received any attention: the differentiation of modifications and phrase schemata, the valence spectrum of phrasemes as well as formulaic expressions with unique components.

---

## 1 Einleitung

Die (unzureichende) Erfassung fester Wortverbindungen in allgemeinsprachigen Wörterbüchern und phraseologischen Spezialwörterbüchern ist bereits mehrfach Gegenstand innerhalb der Phraseologieforschung gewesen (siehe u. a. Pilz 1987, 1995, 2002; Hallsteinsdóttir 2006; Müller/Kunkel-Razum 2007; Korhonen 2011). Dabei fällt das Urteil über die lexikografische Auszeichnung von Phrasemen in den meisten Fällen nicht positiv aus. Bemängelt wird vor allem die Tatsache, dass die Nennformen häufig nicht dem tatsächlichen Sprachgebrauch entsprechen und die Bedeutungsangaben und pragmatischen Gebrauchsbedingungen verkürzt oder überhaupt nicht dargestellt werden (vgl. u. a. Belica/Steyer 2008: 8). Trotz einiger Verbesserungen in den letzten Jahren kann die Behandlung von Phrasemen insbesondere in allgemeinsprachigen Wörterbüchern als wenig befriedigend bezeichnet werden, insofern kein Werk existiert, das den geforderten Ansprüchen voll und ganz entspricht, wie Burger (2015: 182) explizit hervorhebt:

Die Phraseologie-Forschung hat eine ganze Reihe von Forderungen an die Lexikographie formuliert, die zu einer erheblichen Verbesserung der Wörterbücher sowohl hinsichtlich der Makrostruktur als auch der Mikrostruktur (d. h. der Struktur des einzelnen Artikels) führen könnten. Aber es gibt bisher noch kein einziges Wörterbuch, das diese Erkenntnisse und Forderungen in vollem Umfang berücksichtigt hätte.

In Bezug auf phraseologische Spezialwörterbücher hält Burger (2015: 182) fest, dass diese „früher noch schlechter als die einsprachigen allgemeinen Wörterbücher“ gewesen sind, die Situation sich aber deutlich gebessert habe. Als das „Standardwerk unter den deutschen phra-

seologischen Wörterbüchern“ (Heine 2012: 171) kann der Duden 11 „Redewendungen“ (im Jahr 2013 in der vierten Auflage erschienen) genannt werden (siehe u. a. Scholze-Stubenrecht 2004). Zwar wurden im Laufe der Zeit die jeweiligen Auflagen zum Teil verbessert, die Einträge zu den einzelnen Phrasemen im Duden (2013) können aber keineswegs durchgängig als zufriedenstellend bezeichnet werden. Dies ist darauf zurückzuführen, dass auch dieses Nachschlagewerk nicht auf der Auswertung von großen Textsammlungen und somit des wirklichen Sprachgebrauchs basiert, was in einem offensichtlichen Gegensatz zur starken Ausbreitung korpuslinguistischer Methodik innerhalb der Phraseologie steht (vgl. Heine 2012: 176–180).<sup>1</sup>

Der vorliegende Beitrag möchte einen Einblick geben, inwiefern Korpusanalysen für die lexikografische Erfassung von Phrasemen dienlich sein können. Zunächst werden die Problemfelder der bisherigen Phraseografie sowie der Nutzen korpusbasierter Ansätze zusammengefasst. Im Mittelpunkt stehen daraufhin Beispielanalysen, die anhand von drei unterschiedlichen Beschreibungskategorien die Vorteile der Korpusanalyse aufzeigen. Ich greife hierfür Phänomene heraus, denen innerhalb der Phraseografie bislang kaum bzw. keine Aufmerksamkeit geschenkt wurde: die Abgrenzung zwischen Modifikationen und Modellbildungen, das Valenzspektrum von Phrasemen sowie das Phänomen der unikalenen Komponenten.

## 2 Das Hauptproblem der traditionellen Phraseografie: Ausblenden des tatsächlichen Sprachgebrauchs

Die lexikografische Beschreibung fester Wortverbindungen stellt sicherlich eines der problematischsten Aufgabengebiete der Phraseologie dar. Sie bereitet den Lexikografen seit Anbeginn der Phraseologieforschung große Schwierigkeiten. Die meisten phraseologischen Nachschlagewerke weisen folglich mehr oder weniger große Schwächen auf. Neben allgemeinen Mängeln, die unter anderem die verwendete Terminologie, die Aufnahme bzw. Nicht-Aufnahme bestimmter Phrasemklassen, die Darbietungsform, das Ordnungsprinzip sowie den Artikelaufbau der einzelnen Lemmata betreffen (vgl. u. a. Pilz 1995: 309–316), resultieren speziellere Schwierigkeiten vor allem aus der Tatsache, dass in den meisten phraseologischen Lexika die Phraseme nicht auf der Grundlage ihres tatsächlichen Gebrauchs ausgezeichnet werden. Die gängige Praxis sieht vielmehr so aus, dass

die Entscheidung für die Aufnahme in ein Wörterbuch stattdessen zumeist auf der Basis bereits vorhandener Idiomsammlungen und Nachschlagewerke getroffen wird. Dies führt in vielen Fällen zu einer Tradierung bestimmter phraseologischer Kernbestände, die nicht mehr dem aktuellen Usus entsprechen.

(Steyer/Lauer/Brunner 2008: 108)

Das größte Dilemma zeigt sich laut Müller/Kunkel-Razum (2007: 944) in der Diskrepanz zwischen dem aktuellen Stand der Forschung und der „veralteten“ Darstellung in den Wörterbüchern:

---

<sup>1</sup> Heine (2012: 171) führt in Bezug auf Duden 11 an: „Dennoch kann hier nur eingeschränkt von der Arbeit mit Korpora gesprochen werden, da die Belege lediglich dazu dienen, die Verwendung der Phraseologismen zu illustrieren, es sich also bestenfalls um das Konsultationsparadigma [...] handelt, eher aber um eine reine Belegsammlung.“

Für die Phraseologieforscher und Wörterbuchkritiker bleibt damit der Stand der Aufnahme von Phraseologismen in die Wörterbücher und ihre Darbietung dort weit hinter den Ergebnissen ihrer Forschung zurück.

Die entscheidende lexikografische Frage lautet dabei: Wie soll das Lemma eines Phrasems im Wörterbuch aussehen und vor allem auf welchen empirischen Daten sollen die einzelnen Angaben basieren? Sich hierbei am tatsächlichen Sprachgebrauch zu orientieren, steht meines Erachtens außer Frage. Dass dies in der traditionellen Phraseografie keineswegs immer im Vordergrund stand, verdeutlicht Steyer (2010: 255):

Die teils überholte und veraltete Belegsituation in Wörterbüchern des Deutschen oder im Kontrast mit dem Deutschen sowie in Materialien für die Vermittlung des Deutschen wird mittlerweile allgemein als Desiderat angesehen. So haben auch wir häufig die wenig bis gar nicht am aktuellen Usus orientierte Vererbung identischer Beispiele in Phraseologie und Lexikografie kritisiert.

Neuere Projekte wie beispielsweise das Projekt „Usuelle Wortverbindungen“<sup>2</sup> und das „Sprichwortprojekt“<sup>3</sup> des IDS steuern daher bewusst dieser Traditionslinie entgegen, indem die dort zu findenden Artikel strikt auf der Grundlage von Korpusdaten erstellt sind (siehe Steyer/Lauer/Brunner 2008 und Steyer 2012).<sup>4</sup> Sie können als Vorreiter und Vorbilder einer korpusbasierten Phraseografie angesehen werden.

### **3 Der entscheidende Vorteil korpusbasierter Ansätze: Analyse des tatsächlichen Sprachgebrauchs**

Unter korpusbasierter Phraseografie verstehe ich in Anlehnung an Steyer (2010: 249) die lexikografische Erfassung von festen Mehrworteinheiten „durch die Analyse großer Datenmen-gen“. Die Korpusanalyse „eröffnet den empirischen Zugang zu sprachlichem Wissen“ (Steyer 2002: 218) und erlaubt, „sprachlichen Usus in einer völlig neuen Dimension beschreiben zu können“ (Steyer 2003: 45):

Korpusmethodische Zugänge ermöglichen, Wörterbücher zu erarbeiten, die sich am aktuellen Sprachgebrauch orientieren, was immer wieder gefordert wurde und wird.

(Steyer 2013: 292)

Und auch für Dobrovol'skij (2015: 275f.) liegt die Stärke der auf Korpusdaten basierenden Phraseografie darin, die Nennform einer Mehrworteinheit anhand ihres wirklichen Sprachgebrauchs feststellen und lemmatisieren zu können:

[T]he appearance of large text corpora opens up entirely new possibilities for presenting phraseology in dictionaries. For example, registering a phraseme in a dictionary requires selecting its most frequent standard form. Dictionaries have often inherited the form of the lemma from older reference sources. The use of corpora allows us to identify the standard form of a lemma from the point of view of modern usage.

Phraseografie korpusbasiert zu betreiben, bringt demnach den Vorteil mit sich, „authentische Sprachdaten ins Zentrum der lexikographischen Beschreibung rücken zu können“ (Steyer

<sup>2</sup> <http://wvonline.ids-mannheim.de/> (Stand 06.03.2018).

<sup>3</sup> <http://www.spruchwort-plattform.org/> (Stand 06.03.2018).

<sup>4</sup> Für eine umfassendere Übersicht über elektronische Wörterbücher siehe Steyer (2013: Kap. 5.3).

2013: 296), d. h. die tatsächliche Verwendung der entsprechenden Wortverbindungen auf der Grundlage sprachlicher Massendaten erforschen und später in den Lemmata adäquat erfassen zu können. Darunter fallen unter anderem

1. die Ermittlung der Nennform des Phrasems, z. B. in Bezug auf usuelle Varianten, okkasionelle Modifikationen, (obligatorische sowie fakultative) Valenzstellen und lexikalisch offene Phraseme (Modellbildungen)
2. die Ermittlung von Bedeutungsangaben aus korpus-pragmatischer Sicht
3. die Ermittlung pragmatischer Besonderheiten (wie z. B. Sprechaktgebundenheit bestimmter Phraseme)<sup>5</sup>

Es ist an dieser Stelle hervorzuheben, dass die korpusbasierte Phraseografie methodisch gesehen ein pragmatischer Ansatz *par excellence* ist. So betont Steyer (2009: 120), dass die

[k]orpusgesteuerte qualitative Sprachbeschreibung [...] primär pragmatisch [ist]. [...] Funktionales wird aus der Nische subjektiver Perspektiven auf einzelne sprachliche Akte herausgeholt und durch die Möglichkeit, massenhafte Gebrauchsmuster zu erkennen, „kollektiviert“ und damit besser generalisierbar.

Ich schließe ich mich den Ansätzen einer „korpusbasierten Phraseografie“ nach Steyer (2010) an und möchte anhand von ausgewählten Phänomenen aufzeigen, wie mittels eines sprachgebrauchsorientierten Vorgehens die Lemmatisierung fester Mehrworteinheiten verbessert werden kann. Dabei werden im Folgenden drei Phänomene herausgegriffen, denen meines Erachtens bislang eher weniger Beachtung geschenkt wurde, deren korpusbasierte Erforschung aber sowohl einen Beitrag zur Phraseografie als auch zu theoretischen Fragen der Phraseologie allgemein leisten kann. Die Korpusanalysen basieren auf dem Deutschen Referenzkorpus (DeReKo) und dem dazugehörigen Analysesystem COSMAS-II.<sup>6</sup> Als Grundlage der Beobachtungen dient der Abgleich mit Einträgen aus dem Duden (2013).

## 4 Beispielanalysen

### 4.1 Abgrenzung zwischen Modifikationen und Modellbildungen

Der ersten Beispielanalyse werden die folgenden drei Textstellen vorangestellt, die aus dem DeReKo entnommen sind (Suchanfrage: (noch /+w1 "oder") /+w4 schon):

Bildschirm oder Papier? Am Kulturgut Buch scheiden sich die Geister. Dabei geht es nicht um Inhalte, sondern leidenschaftlich wird diskutiert, wie man es sich einverleiben soll, das Buch.  
**Liest du noch oder kindlest du schon?** (St. Galler Tagblatt, 26.11.2012)

**„Trinkst du noch, oder säufst du schon?“** Die Grenzen zwischen unschädlichem Alkoholgenuss und Sucht sind fließend. Die gute Nachricht ist: „Man kann zu jedem Zeitpunkt etwas dagegen tun“, sagt Suchtexperte Wolf Dietrich Braunwarth. (Nürnberger Nachrichten, 01.05.2013)

<sup>5</sup> Siehe hierzu die Ausführungen in Burger (2015: 198f.).

<sup>6</sup> Den Auswertungen liegt das komplette öffentlich zugängliche Archiv W (Hauptarchiv) zugrunde, das „mit über 29 Milliarden Wörtern (Stand 31.03.2016) die weltweit größte linguistisch motivierte Sammlung elektronischer Korpora mit geschriebenen deutschsprachigen Texten aus der Gegenwart und der neueren Vergangenheit“ (<http://www1.ids-mannheim.de/kl/projekte/korpora/> Stand 22.12.2016) darstellt.

Im dichten Schneetreiben wurden Plakate geschwenkt: „**Arbeitest du noch oder bettelst du schon?**“ Das Motto einer kleinen, aber wichtigen neuen Streikgruppe unter den Demonstranten. (Hamburger Morgenpost, 11.03.2006)

Auf den ersten Blick würde man aus phraseologischer Sicht diese Belege wahrscheinlich als Modifikationen des festen Werbeslogans *Wohnst du noch oder lebst du schon?* klassifizieren. Die Verben *wohnen* und *leben* werden substituiert. Es stellt sich aber die Frage, ob es sich bei dieser formelhaften Wendung nicht vielmehr um ein lexikalisch offenes Phrasem handelt, also um eine Modellbildung, in der die Verbalstellen mehr oder weniger frei besetzt werden können (siehe Stumpf 2016).

Um diese Frage beantworten zu können, bietet sich die sogenannte „automatische Slotanalyse“ (Steyer 2013: 112) an.<sup>7</sup> Diese ermöglicht „die automatische Zählung der lexikalischen Füller einer Leerstelle“ (Steyer 2012: 305) sowie „die Ermittlung ihrer Vorkommensproportionen innerhalb eines Slots“ (Steyer 2013: 122). Die Ergebnisse können in sogenannten „Lückenfüllertabellen“ (Steyer 2014: 120) eingesehen werden. Beispielsweise legt Steyer (2013: 126) dar, dass der Slot in dem Phrasem *wie X<sub>[Nomen]</sub> in den Ohren* neben zahlreichen okkasionellen Einmalfüllern zwei lexikalisch dominante Füller (*Musik* und *Hohn*) besitzt, die als feste Nennform existieren. Die Slotanalyse zeigt, dass *Musik* 257mal und *Hohn* 122mal die Leerstelle ausfüllen; alle übrigen Substantive (wie *Spott*, *Ironie* oder *Watte*) sind nur vereinzelt dokumentiert (vgl. Steyer 2013: 124f.) und demnach als Modifikationen zu beurteilen. Die Slotanalyse und die dazugehörigen Lückenfüllertabellen stellen für die Unterscheidung zwischen Modifikationen und Modellbildungen aus dem Grund ein hilfreiches Instrumentarium dar, da sie empirische Hinweise „auf die Natur der lexikalischen Füller und die Skala der Typizität bzw. Produktivität“ (Steyer 2011: 230) und demzufolge auch auf den Verfestigungsgrad bzw. den Leerstellencharakter einer formelhaften Wendung geben können (vgl. Steyer/Brunner 2014: 87).

Die Slotanalyse zeigt, dass sich das geflügelte Wort *Wohnst du noch oder lebst du schon?* mittlerweile zu einer Modellbildung entwickelt hat. Wie anhand der Lückenfüllertabelle zu sehen ist, tritt der eigentliche Werbespruch von IKEA „nur“ in ca. 17% aller Belege auf:

Lückenfüller	Anzahl	Prozentanteil	Lückenfüller	Anzahl	Prozentanteil
<i>Wohnst ... lebst</i>	84	17,07	<i>Sanierst ... reißt</i>	2	0,41
<i>Lachst ... denkst</i>	19	3,86	<i>Weinst ... lachst</i>	2	0,41
<i>Rennst ... lebst</i>	12	2,44	<i>Liest ... tauschst</i>	2	0,41
<i>Lebst ... wohnst</i>	10	2,03	<i>Heulst ... joggst</i>	2	0,41
<i>Sprichst ... kommunizierst</i>	10	2,03	<i>Übst ... tanzt</i>	2	0,41

<sup>7</sup> Das Vorgehen sieht dabei wie folgt aus: Die KWIC-Belege einer COSMAS-II-Suchanfrage, die sowohl aus den öffentlichen als auch nicht-öffentlichen Korpora stammen, werden exportiert und in das Slotanalyse-Tool des Projekts „Usuelle Wortverbindungen“ eingelesen. Im Tool folgt eine Muster-Suchanfrage mit Lücken über die KWIC-Belege. Diese ist rein oberflächenbasiert und erlaubt keine lemmatisierten oder wortartenspezifisierten Anfragen. Beachtet werden sollte, dass die Treffermengen der Muster-Suchanfrage immer geringer sind als die der COSMAS-II-Suchanfrage, da sich die Muster-Suchanfrage stärker auf ein bestimmtes Oberflächenmuster fokussiert. Die Prozentangaben der Lückenfüllertabellen bilden die Treffermengen der Muster-Suchanfrage ab. (Ich danke Annelen Brunner für diese Anmerkungen zum allgemeinen Vorgehen einer Slotanalyse. Zudem bedanke ich mich recht herzlich bei Kathrin Steyer und Annelen Brunner, die mir die Daten der Slotanalysen und -tabellen zur Verfügung gestellt haben. Alle Suchanfragen erfolgten im September 2015.).

<i>Glaubst ... denkst</i>	7	1,42	<i>lebst ... herbst</i>	2	0,41
<i>Suchst ... wohnst</i>	7	1,42	<i>Liest ... kindlest</i>	2	0,41
<i>Schläfst ... träumst</i>	5	1,02	<i>Wohnst ... liest</i>	2	0,41
<i>Lernst ... lebst</i>	5	1,02	<i>Schläfst ... lebst</i>	2	0,41
<i>Träumst ... glaubst</i>	5	1,02	<i>Badest ... duschst</i>	2	0,41
<i>Träumst ... lebst</i>	5	1,02	<i>Küsst ... beißt</i>	2	0,41
<i>Tanzst ... klebst</i>	4	0,81	<i>Lernst ... sparst</i>	2	0,41
<i>Lebst ... herbst</i>	4	0,81	<i>Sparst ... wohnst</i>	2	0,41
<i>Bohlst ... klumst</i>	4	0,81	<i>Sparst ... lebst</i>	2	0,41
<i>Schraubst ... lebst</i>	4	0,81	<i>Managst ... führst</i>	2	0,41
<i>Spielst ... bist</i>	4	0,81	<i>Arbeitest ... spielst</i>	2	0,41
<i>Hörst ... verstehst</i>	4	0,81	<i>Lernst ... turnst</i>	2	0,41
<i>Trinkst ... säufst</i>	3	0,61	<i>Shoppst ... denkst</i>	2	0,41
<i>Arbeitest ... bettelst</i>	3	0,61	<i>Lebst ... kaufst</i>	2	0,41
<i>Lauschst ... schläfst</i>	3	0,61	<i>Lebst ... baust</i>	2	0,41
<i>Schraubst ... wohnst</i>	3	0,61	<i>Wohnst ... klagst</i>	2	0,41
<i>Wohnst ... hartzst</i>	3	0,61	<i>Schmierst ... baust</i>	2	0,41
<i>Rauchst ... lebst</i>	3	0,61	<i>Studierst ... lebst</i>	2	0,41
<i>Arbeitest ... lebst</i>	3	0,61	<i>Läufst ... fliegst</i>	2	0,41
<i>Integrierst ... assimilierst</i>	3	0,61	<i>Zahlst ... gestaltest</i>	2	0,41
<i>Fährst ... lebst</i>	3	0,61	<i>lernst ... lebst</i>	2	0,41
<i>Lachst ... swingst</i>	3	0,61	<i>Lebst ... liegst</i>	2	0,41
<i>muckst ... mackst</i>	2	0,41	<i>Zahlst ... schläfst</i>	2	0,41
<i>Wohnst ... teilst</i>	2	0,41	<i>Stehst ... läufst</i>	2	0,41
<i>Suchst ... lebst</i>	2	0,41	<i>Mietest ... besitzt</i>	2	0,41
<i>Nutzt ... störst</i>	2	0,41	<i>Bügelst ... lebst</i>	2	0,41
<i>Wirst ... bist</i>	2	0,41	<i>Träumst ... planst</i>	2	0,41

**Tabelle 1: (Ausschnitt der) Lückenfüllertabelle zu  $X_{[Verb]} Y_{[Pronomen]}$  *noch* oder  $Z_{[Verb]} Y_{[Pronomen]}$  *schon*?<sup>8</sup>**

Die zahlreichen unterschiedlichen Belege lassen sich somit nicht als Modifikationen, sondern als konkrete Auffüllungen einer zugrunde liegenden Modellbildung interpretieren. Lediglich die grammatische Verknüpfung der zweigliedrigen Struktur in Form der Konjunktion *oder* sowie die beiden Adverbien *noch* und *schon* sind lexikalisch fest. Die übrigen Komponenten, also die beiden Verbstellen und die Pronominalstellen, können mehr oder weniger beliebig besetzt werden, wobei in den meisten Fällen das Pronomen *du* beibehalten wird.<sup>9</sup> Das abstrakte Muster lässt sich wie folgt darstellen:  $X_{[Verb]} Y_{[Pronomen]}$  *noch* oder  $Z_{[Verb]} Y_{[Pronomen]}$  *schon*? Aus semantischer Perspektive besitzt die zweite Verbstelle in der Regel eine positivere Konnotation als die erste, mit der sie in gewisser Weise in einem Gegensatz steht.<sup>10</sup>

Die hier angeführte Wendung und weitere Beispiele (vgl. Stumpf 2016: 329–336) verdeutlichen, dass es bei manchen festen Wortverbindungen zu einem Verschwimmen der Grenze zwischen Modifikation und Modellbildung kommen kann. Nicht bei jedem (veränderten)

<sup>8</sup> In Tabelle 1 sind die ersten 64 Treffer von insgesamt 260 abgebildet.

<sup>9</sup> Bezüglich der Besetzung der Verbstellen lässt sich allerdings anführen, dass in den konkreten Beispielen fast ausschließlich morphologisch recht kurze, ein- und zweiwertige Handlungsverben realisiert sind. Für diese Beobachtung danke ich Sven Staffeldt.

<sup>10</sup> Verwiesen sei hierbei auch auf den entsprechenden Eintrag im OWID-Online-Lexikon zu diesem Phrasem, in dem es heißt: „Dieses Sloganismuster kommt in den Korpora häufiger vor als der Slogan selbst und weist ein hohes Variationsspektrum auf.“ (<http://www.owid.de/artikel/404090>, Stand 06.03.2018).

Phrasem lässt sich also exakt bestimmen, ob es sich bei empirisch belegbaren Varianten um sprachspielerische Veränderungen einer lexikalisch festen Nennform oder vielmehr bereits um eine Wendung mit entsprechender Leerstelle handelt.

Für die lexikografische Erfassung solcher Wendungen ist es wichtig, die Leerstellenstruktur ebenfalls mit zu berücksichtigen und gegebenenfalls statt einer lexikalisch voll realisierten Nennform eine Nennform mit Leerstellen anzugeben (siehe Steyer 2015), wobei die Nicht-Berücksichtigung von Modellbildungen generell als ein Mangel der lexikografischen Beschreibung von Phrasemen angesehen werden kann.<sup>11</sup> In der neuesten Auflage des Dudens (2013) „Redewendungen“ ist jedoch noch nicht einmal das geflügelte Wort selbst erfasst.<sup>12</sup>

## 4.2 Valenzspektrum von Phrasemen

Die phraseografische Erfassung der Valenz von Idiomen kann als unbefriedigend bezeichnet werden (vgl. Stumpf 2015a: 20–22). Mängel treten dabei vor allem in der Darstellung der externen Valenzstellen auf. Diese Problematik steht in einem engen Zusammenhang mit den Abgrenzungsschwierigkeiten zwischen fakultativen und obligatorischen Ergänzungen (vgl. Kühn 2003: 106). So ist es häufig der Fall, dass wichtige obligatorische Ergänzungen in der Nennform des Wörterbuchs nicht verzeichnet sind (vgl. Keil 1997: 63). Denn in der herkömmlichen Kodifizierung findet die (wendungsexterne) Subjektergänzung keine Berücksichtigung.<sup>13</sup> Beispielsweise ist das Phrasem *jmd. schiebt jmdm. etw. in die Schuhe* im Duden (2013: 669) in der Nennform *jmdm. etw. in die Schuhe schieben* verbucht. Diese Darstellungsweise geht auf die Tradition der Wörterbücher zurück, bei Verben den Infinitiv anzugeben, was jedoch dazu führt, dass durch diese Art der Abbildung offensichtlich eine obligatorische Ergänzung ignoriert wird (vgl. Burger 2015: 191 sowie Kjaer 1987: 167).<sup>14</sup>

Eine möglichst vollständige Abbildung des quantitativen und qualitativen Aktantenpotenzials der zu kodifizierenden Phraseme ist dabei nur zu erreichen, wenn die Phraseografie auf umfangreiche Korpusanalysen zurückgreift. Durch die Analyse des tatsächlichen Sprachgebrauchs mithilfe eines korpusbasierten Vorgehens können die intern fixierten Komponenten<sup>15</sup>, die quantitativen und qualitativen externen Valenzstellen sowie die obligatorischen und fakultativen Ergänzungen von Phrasemen aufgedeckt werden.

Zwei Beispiele sollen hier genügen: Die syntaktische Strukturierung der Leerstellen von Phrasemen geht in vielen Fällen über kasuelle Ergänzungen hinaus und erstreckt sich auch über nicht kasuelle Anschlüsse wie beispielsweise Situativ- oder Modifikativergänzungen, die in phraseologischen Wörterbüchern jedoch häufig keine Beachtung finden. So kommt mithilfe einer Korpusanalyse zum Vorschein, dass das verbale Idiom *Fuß fassen* neben einer Ergän-

<sup>11</sup> So finden Modellbildungen bislang kaum Eingang in phraseologische Spezialwörterbücher.

<sup>12</sup> Dies liegt wohl auch daran, dass dieses ein relativ neues Phrasem darstellt und dass geflügelte Worte kaum Berücksichtigung im Duden (2013) finden.

<sup>13</sup> Ergänzungsklassen nach Engel (2009: 134).

<sup>14</sup> Einen Verbesserungsvorschlag liefert unter anderem Wotjak (1985/86, 1986, 1992) mit ihrem mehrstufigen und integrativen Beschreibungsmodell.

<sup>15</sup> In der Forschung spricht man hierbei auch häufig von „internen Valenzstellen“. Da dieser Terminus meines Erachtens aber nicht treffend und ein Widerspruch in sich ist, verwende ich die in Stumpf (2015a: 7) vorgeschlagene Alternative „intern fixierte Komponente“.

zung im Nominativ auch eine Situativergänzung fordert, die im Duden (2013: 244) nicht registriert ist:

Der Osteroder Landrat Bernhard Reuter (SPD) warnte vor den wirtschaftlichen und sozialen Folgen, wenn die Rechtsextremisten im Südharz stärker **Fuß fassen** sollten: „Wer möchte Urlaub machen in einer Gegend, in der sich Nazis breit machen?“ (Hannoversche Allgemeine, 28.09.2007)

Mark versucht, an der Realschule **Fuß zu fassen**. Seine Noten hat er stabilisiert, „paar Dreier, mehrere Vieren“, meint er. (Braunschweiger Zeitung, 27.01.2006)

Die heimlichen Hersteller **fassen** auch auf dem alten Kontinent **Fuß**: „Europa hat spät angefangen, holt aber rasend schnell auf“, bestaunt Flextronics-Manager Mike McNamara die Auslagerungsaktivitäten hiesiger High-Tech-Giganten. (FOCUS, 05.02.2001)

Ein weiteres Phrasem mit einer im Duden (2013: 267) nicht verzeichneten Ergänzung ist *mit jmdm. ins Gericht gehen*. Korpusabfragen verdeutlichen, dass das Idiom eine Modifikativergänzung fordert, die genauer beschreibt, wie – auf welche Art und Weise – mit jemandem „ins Gericht gegangen wird“:

Kunert wollte mit seinem Team trotzdem nicht allzu hart **ins Gericht gehen**. „Nach der Pause kann ich meiner Mannschaft keinen Vorwurf machen. Da spielten wir guten und druckvollen Fußball. Daraus resultierten zahlreiche Möglichkeiten“, sagte der Coach. (Braunschweiger Zeitung, 18.03.2013)

Eberhard Diepgen **geht** indes wieder einmal scharf mit seinem fast chancenlosen SPD-Kontrahenten Momper **ins Gericht**. (Berliner Morgenpost, 15.09.1999)

Kritisch **geht** die Kreistagsfraktion auch mit dem Vorgehen des Landrats **ins Gericht**, der ein Votum des Kreistagspräsidiums genutzt habe, um nach außen den Eindruck zu erwecken, der Umzug nach Heusenstaamm sei bereits gelaufen. (Frankfurter Rundschau, 11.12.1998)

Das Phrasem ist zum Zeitpunkt der Abfrage 8.302-mal belegt. Die Besetzung der modalen Leerstelle ist dabei lexikalisch eingeschränkt. Die Korpusanalyse zeigt, dass sie am häufigsten mit *hart* ausgefüllt wird (siehe Tabelle 2). Die Tabelle verdeutlicht aber auch, dass es sich bei den lexikalischen Füllern um Elemente aus einer relativ geschlossenen Gruppe mit einer semantischen Invariante, also um (partiell) synonyme Adjektive handelt. Die Adjektive besitzen alle eine intensivierende/bewertende Bedeutung, mit der die Art und Weise beschrieben wird, wie mit jemandem „ins Gericht gegangen“ wird.

Modifikativergänzung	Trefferzahl im DeReKo
<i>hart</i>	5.585
<i>scharf</i>	769
<i>kritisch</i>	212
<i>streng</i>	135
<i>heftig</i>	82
<i>schonungslos</i>	53
<i>harsch</i>	45
<i>gnadenlos</i>	24

Tabelle 2: Quantitative Auswertung der Modifikativergänzung des Idioms  
*jmd. geht mit jmdm. ins Gericht*



Das Beispiel veranschaulicht, dass es bei manchen Phrasemen schwierig ist, zwischen den bereits in der Nennform des Phrasems fixierten Komponenten und den lexikalisch offenen externen Valenzstellen exakt zu unterscheiden. So könnte es sich bei der modalen Erweiterung aufgrund ihrer relativ breiten Variation um eine (variable) externe Valenz handeln (*jmd. geht mit jmdm. auf eine intensive/bewertende Art und Weise ins Gericht*). Wahrscheinlicher ist jedoch, dass das Adjektiv *hart* aufgrund seiner hohen Frequenz fest in der Nennform des Idioms *jmd. geht mit jmdm. hart ins Gericht* verankert ist. Die alternativen Auffüllungen wie zum Beispiel *scharf*, *kritisch* oder *streng* sind dann als Varianten bzw. Modifikationen zu interpretieren. So oder so: Diese Ergänzung sollte auf jeden Fall in den Wörterbucheinträgen verbucht werden – ob als fester Bestandteil oder als Leerstelle.

### 4.3 Unikale Komponenten

Als abschließendes Beispiel wird im Folgenden das Phänomen der sogenannten unikalen Komponenten bzw. Unikalia thematisiert. Unter Unikalia versteht man in der Phraseologieforschung Wörter, die nur (noch) im Konstituentenbestand von Phrasemen auftreten, wie beispielsweise *klipp* und *klar*, *etw. auf dem Kerbholz haben* oder *frech sein/schimpfen wie ein Rohrspatz*. Korpusauswertungen des tatsächlichen Vorkommens vermeintlich unikaler Komponenten zeigen jedoch, dass es sich hierbei um eine prototypische Kategorie handelt (vgl. Stumpf 2014: 101–108, 2018: 71–74). Unikalia sind nicht „entweder – oder“, sondern „mehr oder weniger“ phraseologisch gebunden, weshalb Unikalität aus korpuslinguistischer Sicht als eine graduelle Eigenschaft betrachtet werden muss:

Neben Wörtern, die ausschließlich in einem formelhaften Kontext realisiert sind (z. B. *Schnippchen* in *jmdm. ein Schnippchen schlagen*), fördert die Korpusanalyse auch solche Wörter zutage, die nur in geringem Maße gebunden sind (z. B. *Gardinenpredigt* in *jmdm. eine Gardinenpredigt halten*). Das eigentlich Interessante zeigt sich jedoch darin, dass zahlreiche Wörter im Zwischenbereich dieser beiden Extrempunkte anzusiedeln sind (z. B. *Schokoladenseite* und *Armutszugnis*). Korpusanalytisch lässt sich also eine graduelle Verteilung von stark phraseologisch bis hin zu kaum phraseologisch gebundenen Konstituenten feststellen.

(Stumpf 2015b: 98)<sup>16</sup>

Korpusanalysen und das von Steyer (2013) entwickelte Konzept der usuellen Wortverbindungen erweisen sich als äußerst hilfreich, wenn es um die Frage geht, welche Komponenten unikaler Natur sind. Darüber hinaus hilft ein korpusbasiertes Vorgehen generell bei der Frage, ob es sich bei bestimmten Wortverbindungen überhaupt um formelhafte Wendungen handelt, in die Unikalia eingebettet sind, zum Beispiel bei der Wortverbindung *am/an den Stadtrand*. Wertet man diese korpuslinguistisch aus, so zeigt sich, dass hier eine formelhafte Wendung mit mehr oder weniger gebundener Komponente vorliegt. Denn im DeReKo tritt das Wort *Stadtrand* in über 80% der Belege mit der Präposition *an/am* auf, was auf eine usuell gewordene, sprich feste Wortverbindung hindeutet:

Bedenklich sei auch, dass auf diese Weise viele innenstadtrelevante Waren **an den Stadtrand** kämen. Das könnte nicht nur den Händlern deutlich schaden, sondern die erfolgreichen Bemühungen der Stadt zur Belebung der Innenstadt zunichte machen. (Nordkurier, 11.04.2015)

<sup>16</sup> Die graduelle Verteilung der phraseologischen Gebundenheit von insgesamt 1.318 Komponenten ist in Stumpf (2015b: 479–525) einsehbar.

Wohnungen in der Innenstadt statt Container **am Stadtrand**: Architekturstudenten der Leibniz Universität in Hannover haben alternative Wohnmodelle für Flüchtlinge entwickelt. (die tageszeitung, 04.02.2015)

Solche distributionell eingeschränkten Wörter können demnach aus korpusbasierter Perspektive als mehr oder weniger „unikal“ klassifiziert werden, auch wenn die Wortverbindungen, in die sie eingebettet sind, auf den ersten Blick nicht im klassischen Sinne phraseologisch erscheinen. Aus korpuslinguistischer Sicht lässt sich also sagen: Ist die syntagmatische Verketzung eines Wortes in hohem Maße auf nur (noch) sehr wenige, bestimmte Lexeme eingeschränkt, liegt eine (mehr oder weniger) phraseologisch gebundene Komponente und demnach auch eine formelhafte Wendung vor, in der diese realisiert ist.<sup>17</sup>

Im Zuge einer korpusbasierten Analyse kommen somit usuelle Wortverbindungen mit unikal Komponenten zum Vorschein, die bisher seitens der Phraseologieforschung nicht beachtet wurden und wahrscheinlich auch niemals als phraseologisch bzw. formelhaft identifiziert worden wären. In solchen Fällen handelt es sich nicht mehr um die klassischen, idiomatischen und oft als „irregulär“ bezeichneten Unikalia-Phraseme, sondern um Wörter, die sich durch Verbindungskonventionen mit anderen Wörtern, durch Traditionen des Formulierens und somit durch Gebrauchskonventionen zu unikal Komponenten entwickelt haben oder sich gerade auf dem Weg zur Unikalität befinden (vgl. Steyer 2004: 91). In Tabelle 3 sind einige solcher Fügungen und ihre formelhafte Gebundenheit aufgelistet.<sup>18</sup> Es ist zu erkennen, dass es sich hierbei größtenteils um präpositionale Verfestigungen oder um Nomen-Verb-Kollokationen handelt.

Unikalia-Kandidat	verfestigte Einbettung/formelhafte Wendung	Gebundenheit
<i>Eigendynamik</i>	eine <b>Eigendynamik</b> gewinnen/entwickeln/entstehen/bekommen/entfalten	54%
<i>Lebensunterhalt</i>	seinen <b>Lebensunterhalt</b> von/mit etw. bestreiten/finanzieren/sichern // seinen <b>Lebensunterhalt</b> verdienen	60%
<i>Sekundenbruchteil</i>	innerhalb von/in/innert/binnen/für <b>Sekundenbruchteile(n)</b>	63%

<sup>17</sup> Über den phraseologischen Status einiger Wortverbindungen, die meiner Untersuchung (Stumpf 2015b) zugrunde liegen, kann sicherlich gestritten werden. Es muss daher Folgendes betont werden: Der Unikalia-Auswertung liegt die Annahme zugrunde, dass es in einer Sprache Wörter gibt, die gewisse syntagmatische Beschränkungen aufweisen in dem Sinne, dass diese mehr oder weniger nur in Verbindung mit anderen Wörtern auftreten. Mittels Kookkurrenzanalyse und detaillierter Suchanfragen werden zunächst einmal nur „Wörter“ und nicht „Unikalia“ im Hinblick auf ihre syntagmatische Entfaltung ausgewertet. Mit anderen Worten: Sind die Wörter auf bestimmte „formelhafte“ (Wort-)Umgebungen eingeschränkt? Die Frage, ob es sich bei einer Mehrwortverbindung wie *am/an den Stadtrand* um ein Phrasem handelt bzw. ob *Stadtrand* eine unikale Komponente ist, möchte ich in erster Linie gar nicht beantworten (was angesichts der unscharfen Grenzen zwischen „Phraseologie“ und „freiem“ Sprachgebrauch bzw. zwischen Unikalität und Nicht-Unikalität ohnehin nicht möglich ist). Ich sage also nicht, dass das Wort *Stadtrand* zu so und so viel Prozent unikal (im klassischen Sinne) ist, sondern lediglich, dass dieses zu so und so viel Prozent mit der Präposition *am/an* vorkommt. Das ist ein kleiner, aber feiner Unterschied.

<sup>18</sup> Die Gebundenheit wird in Prozent angegeben, wobei sich die Zahl darauf bezieht, in wie vielen Belegen die Komponente im DeReKo in einer formelhaften Umgebung (d. h. in Verbindung mit bestimmten anderen Wörtern) auftritt. So ist *Zeitlupe* beispielsweise in 65% aller Treffer mit der Präposition *in* realisiert.

<i>Zeitlupe</i>	<i>in <b>Zeitlupe</b></i>	65%
<i>Tagesanbruch</i>	<i>vor/bei/nach <b>Tagesanbruch</b></i>	68%
<i>Unding</i>	<i>etw. ist ein <b>Unding</b></i>	69%
<i>Gesetzeskraft</i>	<i><b>Gesetzeskraft</b> haben/geben/erlangen/erhalten</i>	70%
<i>Dienstschluss</i>	<i>(kurz) vor/nach <b>Dienstschluss</b></i>	75%
<i>Schnäppchenpreis</i>	<i>zum <b>Schnäppchenpreis</b></i>	76%
<i>Abhilfe</i>	<i><b>Abhilfe</b> schaffen/leisten/bringen/sorgen/bieten/verschaffen/versprechen // für <b>Abhilfe</b> sorgen</i>	79%
<i>Zehenspitze</i>	<i>auf (den/die) <b>Zehenspitzen</b></i>	79%
<i>Geschmackssache</i>	<i>etw. ist <b>Geschmackssache</b></i>	81%
<i>Vortag</i>	<i>am/vom <b>Vortag</b></i>	81%
<i>Anschein</i>	<i>den <b>Anschein</b> haben/erwecken/machen // dem/allen <b>Anschein(s)</b> nach</i>	82%
<i>Mordverdacht</i>	<i>unter <b>Mordverdacht</b> (stehen) // wegen <b>Mordverdacht</b> // in <b>Mordverdacht</b> (geraten)</i>	84%
<i>Rettungsschuss</i>	<i>der finale <b>Rettungsschuss</b></i>	89%
<i>Familienkreis</i>	<i>im engsten <b>Familienkreis</b></i>	89%
<i>Augenwinkel</i>	<i>aus dem/im <b>Augenwinkel</b></i>	89%
<i>Aufschluss</i>	<i>über etw. <b>Aufschluss</b> geben // <b>Aufschluss</b> gewinnen/erhoffen/liefern/bringen</i>	89%
<i>Umkreis</i>	<i>im/aus dem <b>Umkreis</b> von X<sub>[Nominalphrase]</sub></i>	94%
<i>Faible</i>	<i>ein <b>Faible</b> für X<sub>[Nominalphrase]</sub> (haben)</i>	95%
<i>Platzgrund</i>	<i>aus <b>Platzgründen</b></i>	98%
<i>Morgenstunde</i>	<i>in/seit/vor/den/zur <b>Morgenstunde</b></i>	98%

**Tabelle 3: Formelhafte Gebundenheit von Wörtern in nicht-idiomatischen, usuellen Wortverbindungen**

Insgesamt zeigen die korpuslinguistischen Auswertungen neben der Tatsache, dass Unikalität eine graduelle Eigenschaft ist, auch, dass zahlreiche formelhaft gebundene Wörter von der bisherigen Forschung nicht erfasst sind, deren Gebundenheit man sich bislang also gar nicht bewusst gewesen ist (z. B. zum *Nulltarif*, meine *Wenigkeit*, *stutzig* machen/werden, aus *Platzgründen* und *am/an den Stadtrand*) (vgl. Stumpf 2015b: Kap. 4). Auch diese sollten selbstverständlich Eingang in phraseologische Wörterbücher finden, eventuell mit dem Hinweis auf die phraseologische Gebundenheit der einzelnen Komponenten.

Wie eine umfassende korpusbasierte Beschreibung eines Phrasems mit unikalener Komponente aussehen könnte, soll mithilfe von Tabelle 4 verdeutlicht werden. Neben Angaben zur Trefferzahl im DeReKo, zur Wortart der unikalener Komponente sowie zur syntaktischen Klassifizierung und zur Bestimmung der phraseologischen Klasse werden eine korpuslinguistisch ermittelte Bedeutungsparaphrase sowie Angaben zur formelhaften Einbettung des Phrasems gemacht und die Prozentzahl festgehalten, in wie vielen Korpusbelegen das Wort in einem formelhaften Kontext auftritt. Es werden darüber hinaus Beispiele für den formelhaften und den außerformelhaften Gebrauch angeführt. Strukturelle, semantische und pragmatische Besonderheiten runden den Eintrag ab.

Komponente	<i>Euphoriebremse</i>
Trefferzahl im DeReKo	1.513 (bei ca. 29 Milliarden Wörtern/Tokens)
Wortart	Substantiv
formelhafte Einbettung/ Nennform und Variationen	<p><i>auf die Euphoriebremse treten</i> (1.008 = 70,5%)  <i>auf die Euphoriebremse drücken</i> (127 = 9%)  <i>die Euphoriebremse treten</i> (112 = 8%)  <i>auf die Euphoriebremse steigen</i> (81 = 5,5%)  <i>der Tritt auf die Euphoriebremse</i> (41 = 3%)  <i>auf der Euphoriebremse stehen</i> (27 = 1,5%)  <i>auf der/die Euphoriebremse</i> (+ Verb) (23 = 1%)  <i>die Euphoriebremse drücken</i> (9 = 0,5%)</p> <p>Insg. 1.428</p>
formelhafte Gebundenheit	94%
syntaktische Klassifizierung nach Burger (2015)	verbales Phrasem / nominales Phrasem ( <i>der Tritt auf die Euphoriebremse</i> )
phraseologische Klasse	Teil-Idiom
Bedeutung des Phrasems	„dafür plädieren, die Erwartungen nicht zu hoch zu schrauben“
Beispiel für formelhaften Gebrauch	<p>Und was ist mit dem VfL in dieser Saison noch möglich? Ein Uefa-Cup-Platz rückt schließlich immer näher. „Für die Fans und den Verein wäre es natürlich geil, nächste Saison international zu spielen“, meint Dejagah, <b>tritt</b> aber zugleich <b>auf die Euphoriebremse</b>: „Es ist wichtig, dass wir auf dem Boden bleiben. Wenn wir so weiterarbeiten, werden wir am Ende auch belohnt.“ (Braunschweiger Zeitung, 07.04.2008)</p>
Beispiel für außerformelhaften Gebrauch	<p>„Es dürfte eigentlich nicht so schwer werden“, meint SG-Betreuer Reiner Schneider, sonst eher ein Freund der <b>Euphoriebremse</b>. Da die Gastgeber in der gesamten Saison nur im Spiel gegen die HSG Kastellaun/Simmern II einen Punkt erspielen konnten, ist selbst Schneider vor dem sonntäglichen Auswärtsspiel optimistisch: „Man kann ja pessimistisch sein, aber eine Niederlage sollte es am Wochenende nicht geben.“ (Rhein-Zeitung, 27.01.1996)</p>

strukturelle, semantische oder pragmatische Besonderheiten	<ul style="list-style-type: none"> <li>• häufig in Verbindung mit einschränkenden Wörtern wie <i>aber</i>, <i>jedoch</i>, <i>hingegen</i> oder mit verstärkenden Wörtern wie <i>kräftig</i>, <i>vehement</i>, <i>heftig</i></li> <li>• häufig im sportlichen Bereich (insb. dem Fußball) zu finden (in Zusammenhang mit Aussagen von Fußballtrainern und -spielern)</li> <li>• steht fast ausschließlich unmittelbar vor wörtlichen Zitaten von Personen und dort als Einschränkung/Relativierung gegenüber einer vorher getätigten Aussage, z. B.: „Sechs Punkte haben wir schon, die nimmt uns keiner mehr weg“, sagte Präsident Michael A. Roth, <b>trat dann aber vorsorglich auf die Euphoriebremse</b>. „Wir dürfen jetzt nicht größenwahnsinnig werden. Ein guter Mittelfeldplatz ist nach wie vor unser Wunschtraum.“ (Mannheimer Morgen, 21.08.2006)</li> </ul>
--	--

Tabelle 4: Exemplarischer korpusbasierter Wörterbucheintrag zur Unikalia *Euphoriebremse*

## 5 Schluss und Ausblick

Möchte man den tatsächlichen Gebrauch von Phrasemen lexikografisch erfassen, so erweist sich ein korpuslinguistisches Vorgehen als effiziente Methode. Die korpusbasierte Phraseografie ist imstande, die genaue Nennform eines Phrasems mit etwaigen Besonderheiten wie zum Beispiel modellartigen Strukturen, internen und externen Valenzstellen sowie unikalen Komponenten aufzudecken.

Für die Zukunft ist es vor allem die Verbindung von korpusbasierter Methodik und online-frei-zugänglicher Präsentation, die die Phraseografie der nächsten Jahre erheblich verbessern kann. Dies betont Holzinger (2013: 65) am Ende eines kritischen Überblicks über die bisherige Phraseografie explizit:

Zur Verbesserung zukünftiger Wörterbuchgenerationen ist v. a. auch eine ständige konsequente Aktualisierung auf der Basis von in Korpusanalysen gewonnenen Einsichten notwendig. Dazu besonders geeignet sind Online-Wörterbücher, die idealerweise sowohl über eine lexikografisch ausgearbeitete Präsentation verfügen als auch Zugang zu einem Korpus gewähren sollten.

Es ist jedoch anzumerken, dass eine solche strikt korpuslinguistische Herangehensweise, wie sie im vorliegenden Artikel vorgestellt wird, einen erheblichen zeitlichen Aufwand bedeutet. Für jede formelhafte Wendung ausführliche Korpusanalysen durchzuführen, die auf strukturelle, semantische und pragmatische Besonderheiten abzielen, wäre sicherlich ein sehr ambitioniertes Vorhaben. So wichtig das Plädoyer für eine korpusbasierte Phraseografie also auch sein mag, in der Praxis ist eine vollständige Umsetzung dessen nur schwer möglich. Es liegt wohl auf der Hand, dass das gesamte phraseologische bzw. vorgeformte Ausdrucksrepertoire einer Sprache auf diese Weise nicht beschrieben werden kann.<sup>19</sup>

<sup>19</sup> Diesbezüglich könnten Ansätze aus dem Bereich der Computerlinguistik/NLP hilfreich sein, mit denen sich bestimmte Suchabfragen automatisieren ließen. Ein intensiverer Austausch zwischen Phraseografie und Computerlinguistik/NLP ist daher wünschenswert, wobei der Frage nachzugehen ist, welche der im Aufsatz beschriebenen Schritte so operationalisierbar sind, dass sie weitgehend maschinell ablaufen können bzw. wo die Interpretation von Seiten des Phraseografen notwendig ist.

Das heißt aber nicht, dass man sich angesichts dessen abschrecken lassen sollte. So könnte man sich beispielsweise zunächst auf bestimmte Klassen an Phrasemen beschränken, um nach und nach den Gesamtbereich des Formelhaften zu lemmatisieren. Das Kollokationenwörterbuch von Annelies Häcki Buhofer u. a. kann dabei als ein positives Beispiel angesehen werden.<sup>20</sup> Aber auch die Auswertungen im Zuge des Projekts „Usuelle Wortverbindungen“ und des „Sprichwörterprojekts“ von Kathrin Steyer sind hierbei hervorzuheben. Zwei Klassen, deren korpuslinguistische Beschreibung in nächster Zeit sicherlich interessant wäre, sind Modellbildungen und Unikalia, wobei hier der Fokus vor allem auf den Bedürfnissen der Wörterbuchbenutzer (z. B. Lerner des Deutschen als Fremd- oder Zweitsprache) liegen sollte.<sup>21</sup>

## Literaturverzeichnis

- Belica, Cyril/Steyer, Kathrin (2008): „Korpusanalytische Zugänge zu sprachlichem Usus“. In: Vachková, Marie (ed.): *Beiträge zur bilingualen Lexikographie*. Prag, Charles University in Prague: 7–24.
- Burger, Harald (2015): *Phraseologie. Eine Einführung am Beispiel des Deutschen*. 5., neu bearbeitete Auflage. Berlin: Erich Schmidt. (= *Grundlagen der Germanistik* 36).
- Dobrovol'skij, Dmitrij (2015): „Introduction“. *International Journal of Lexicography* 3/28: 275–278.
- Duden (2013): *Redewendungen. Wörterbuch der deutschen Idiomatik*. Berlin: Dudenverlag. (= *Duden* 11).
- Engel, Ulrich (2009): *Syntax der deutschen Gegenwartssprache*. 4., völlig neu bearbeitete Auflage. Berlin: Erich Schmidt (= *Grundlagen der Germanistik* 22).
- Hallsteinsdóttir, Erla (2006): „Phraseographie“. *Hermes – Journal of Language and Communication Studies* 36: 91–128.
- Heine, Antje (2012): „Empirische Methoden der Phraseologieforschung im Wandel der Zeit“. In: Prinz, Michael/Richter-Vapaatalo, Ulrike (eds.): *Idiome, Konstruktionen, „verblühte Rede“*. *Beiträge zur Geschichte der germanistischen Phraseologieforschung*. Stuttgart, Hirzel: 165–185. (= *Beiträge zur Geschichte der Germanistik* 3).
- Holzinger, Herbert J. (2013): „Unikale Elemente: Eine Herausforderung für Lexikologie und Lexikografie“. *Aussiger Beiträge: germanistische Schriftenreihe aus Forschung und Lehre* 7: 53–66.
- Keil, Martina (1997): *Wort für Wort. Repräsentation und Verarbeitung verbaler Phraseologismen*. Tübingen: Niemeyer. (= *Sprache und Information* 35).
- Kjaer, Anne Lise (1987): „Zur Darbietung von Phraseologismen in einsprachigen Wörterbüchern des Deutschen aus der Sicht ausländischer Textproduzenten“. In: Korhonen, Jarmo (ed.): *Beiträge zur allgemeinen und germanistischen Phraseologieforschung. Internationales Symposium in Oulu 13.–15.6.1986*. Oulu, Germanistisches Institut: 165–181. (= *Veröffentlichungen des Germanistischen Instituts* 7).

<sup>20</sup> <http://www.kollokationenwoerterbuch.ch/web/> (Stand 22.12.2016).

<sup>21</sup> Zudem sollte man sich in Zukunft genauer Gedanken über die empirische Grundlage der korpusbasierten Phraseografie machen. Dabei geht es vor allem um die Frage der Korpuszusammensetzung – insbesondere im Hinblick auf die Textsortenauswahl. Neben stark konzeptionell geschriebenen Korpora wie das DeReKo wäre es sicherlich nicht von Nachteil, konzeptionell gesprochene Texte bei der phraseografischen Beschreibung zu berücksichtigen (beispielsweise wenn es um die Erstellung eines Lernerwörterbuches geht, in dem Routineformeln und gesprächsspezifische Phraseme im Mittelpunkt stehen sollen).

- Korhonen, Jarmo (ed.) (2011): *Phraseologie und Lexikografie. Phraseologismen in ein- und zweisprachigen Wörterbüchern mit Deutsch*. Burlington, Vermont: Proverbium. (= *Supplement series of Proverbium* 32).
- Kühn, Peter (2003): „Phraseme im Lexikographie-Check. Erfassung und Beschreibung von Phrasemen im einsprachigen Lernwörterbuch“. *Lexicographica* 19: 97–118.
- Müller, Peter O./Kunkel-Razum, Kathrin (2007): „Phraseographie des Deutschen“. In: Burger, Harald u. a. (eds.): *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung*. 2. Halbband. Berlin, de Gruyter: 939–949. (= *Handbücher zur Sprach- und Kommunikationswissenschaft* 28).
- Pilz, Klaus Dieter (1987): „Allgemeine und phraseologische Wörterbücher. Brauchen wir überhaupt phraseologische Wörterbücher?“ In: Korhonen, Jarmo (ed.): *Beiträge zur allgemeinen und germanistischen Phraseologieforschung. Internationales Symposium in Oulu 13.–15. Juni 1986*. Oulu, Germanistisches Institut: 129–153. (= *Veröffentlichungen des Germanistischen Instituts* 7).
- Pilz, Klaus Dieter (1995): „Duden 11. Redewendungen (...). Das anhaltende Elend mit den phraseologischen Wörterbüchern (Phraseolexika)“. In: Baur, Rupprecht S./Chlosta, Christoph (eds.): *Von der Einwortmetapher zur Satzmetapher*. Bochum, Brockmeyer: 305–320. (= *Studien zur Phraseologie und Parömiologie* 6).
- Pilz, Klaus Dieter (2002): „Vorschläge für ein Phraseolexikon der deutschen Sprache oder Vorschläge für ein Lexikon der deutschen Phraseme/Phraseologismen“. In: Hartmann, Dietrich/Wirrer, Jan (eds.): *Wer A sägt, muss auch B sägen. Beiträge zur Phraseologie und Sprichwortforschung aus dem Westfälischen Arbeitskreis*. Baltmannsweiler, Schneider Verlag Hohengehren: 299–311. (= *Phraseologie und Parömiologie* 9).
- Scholze-Stubenrecht, Werner (2004): „Duden 11 – Lexikografisches Konzept und lexikografische Praxis“. In: Steyer, Kathrin (ed.): *Wortverbindungen – mehr oder weniger fest*. Berlin, de Gruyter: 348–359.
- Steyer, Kathrin (2002): „Wenn der Schwanz mit dem Hund wedelt. Zum linguistischen Erklärungspotenzial der korpusbasierten Kookkurrenzanalyse“. In: Hass-Zumkehr, Ulrike/Kallmeyer, Werner/Zifonun, Gisela (eds.): *Ansichten der deutschen Sprache. Festschrift für Gerhard Stickel zum 65. Geburtstag*. Tübingen, Narr: 215–236. (= *Studien zur Deutschen Sprache* 25).
- Steyer, Kathrin (2003): „Korpus, Statistik, Kookkurrenz. Lässt sich Idiomatisches „berechnen“?“ In: Burger, Harald/Häcki Buhofer, Annelies/Gréciano, Gertrud (eds.): *Flut von Texten – Vielfalt der Kulturen. Ascona 2001 zur Methodologie und Kulturspezifität der Phraseologie*. Baltmannsweiler, Schneider Verlag Hohengehren: 33–46. (= *Phraseologie und Parömiologie* 14).
- Steyer, Kathrin (2004): „Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikografische Perspektiven“. In: Steyer, Kathrin (ed.): *Wortverbindungen – mehr oder weniger fest*. Berlin/New York, de Gruyter: 87–116.
- Steyer, Kathrin (2009): „Zwischen theoretischer Modellierung und praxisnaher Anwendung. Zur korpusgesteuerten Beschreibung usueller Wortverbindungen“. In: Mellado Blanco, Carmen (ed.): *Theorie und Praxis der idiomatischen Wörterbücher*. Tübingen, Niemeyer: 119–145. (= *Lexicographica. Series Maior* 135).
- Steyer, Kathrin (2010): „Korpusbasierte Phraseographie. Neue empirische Methoden und Beschreibungsformen“. In: Mellado Blanco, Carmen u. a. (eds.), *La fraseografía del S.*

- XXI. *Nuevas propuestas para el español y el alemán*. Berlin, Frank & Timme: 249–277. (= *Romanistik* 6).
- Steyer, Kathrin (2011): „Von der sprachlichen Oberfläche zum Muster. Zur qualitativen Interpretation syntagmatischer Profile ». *Travaux neuchâtelois de linguistique* 55 : 219–239.
- Steyer, Kathrin (2012): „Sprichwortstatus, Frequenz, Musterbildung. Parömiologische Fragen im Lichte korpusmethodischer Empirie“. In: Steyer, Kathrin (ed.): *Sprichwörter multilingual. Theoretische, empirische und angewandte Aspekte der modernen Parömiologie*. Tübingen, Narr: 287–314. (= *Studien zur Deutschen Sprache* 60).
- Steyer, Kathrin (2013): *Usuelle Wortverbindungen. Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht*. Tübingen, Narr. (= *Studien zur Deutschen Sprache* 65).
- Steyer, Kathrin (2014): „Ohne Vorliegen von Voraussetzungen. Ein historisches Syntagma im Netz produktiver Wortbindungsmuster“. In: Dalmas, Martine/Piirainen, Elisabeth (eds.): *Figurative Sprache – Figurative Language – Langage figuré. Festgabe für Dmitrij O. Dobrovol'skij*. Tübingen, Stauffenburg: 117–130. (= *Stauffenburg Linguistik* 83).
- Steyer, Kathrin (2015): „Patterns. Phraseology in a state of flux“. *International Journal of Lexicography* 3/28: 279–298.
- Steyer, Kathrin/Brunner, Annelen (2014): „Contexts, Patterns, Interrelations – New Ways of Presenting Multi-word Expressions“. *EACL 2014: 14th Conference of the European Chapter of the Association for Computational Linguistics. Proceedings of the 10th Workshop on Multiword Expressions (MWE 2014), 26–27 April 2014*. Gothenburg, Sweden, 2014: 82–88.
- Steyer, Kathrin/Lauer, Meike/Brunner, Annelen (2008): „Usuelle Wortverbindungen in O-WID. Lexikografische Aspekte unter besonderer Berücksichtigung der Pragmatik“. In: Klosa, Annette (ed.): *Lexikografische Portale im Internet*: 107–121. (= *OPAL-Sonderheft* 1/2008).
- Stumpf, Sören (2014): „Mit Fug und Recht – Korpusbasierte Erkenntnisse zu phraseologisch gebundenen Formativen“. *Sprachwissenschaft* 1/39: 85–114.
- Stumpf, Sören (2015a): „Phraseologie und Valenztheorie. Status quo, Forschungsprobleme und (korpusanalytische) Perspektiven“. *Yearbook of Phraseology* 6: 3–34.
- Stumpf, Sören (2015b): *Formelhafte (Ir-)Regularitäten. Korpuslinguistische Befunde und sprachtheoretische Überlegungen*. Frankfurt am Main: Peter Lang. (= *Sprache – System und Tätigkeit* 67).
- Stumpf, Sören (2016): „Modifikation oder Modellbildung? Das ist hier die Frage – Abgrenzungsschwierigkeiten zwischen modifizierten und modellartigen Phrasemen am Beispiel formelhafter (Ir-)Regularitäten“. *Linguistische Berichte* 2/247: 317–342.
- Stumpf, Sören (2018): „Free usage of German unique components. Corpus linguistics, psycholinguistics and lexicographical approaches“. In: Arndt-Lappe, Sabine/Braun, Angelika/Moulin, Claudine/Winter-Froemel, Esme (eds.): *Expanding the lexicon. Linguistic Innovation, Morphological Productivity, and Ludicity*. Berlin/Boston, de Gruyter: 67–89. (= *The Dynamics of Wordplay* 5).
- Wotjak, Barbara (1985/86): „Zu Inhalts- und Ausdrucksstruktur ausgewählter somatischer Phraseolexeme (1/2)“. *Deutsch als Fremdsprache* 4/22: 216–223 u. 5/22: 270–277.
- Wotjak, Barbara (1986): „Zu einer integrativen Mehrebenenbeschreibung von Phraseologismen“. *Deutsch als Fremdsprache* 6/23: 326–331.



Wotjak, Barbara (1992): *Verbale Phraseolexeme in System und Text*. Tübingen: Niemeyer. (= *Reihe germanistische Linguistik* 125).